

## 결합형 양방향 필터를 이용한 깊이 영상의 후처리 필터링 방법

호요성\* · 이상범\*\*

### 1. 서 론

영화 “아바타”의 흥행으로 인해 3차원 영상에 대한 관심이 폭발적으로 증가하면서부터, 이에 따른 파급 효과로 다양한 장르의 3차원 영화가 극장에서 상영되기 시작했고 심지어 3차원 TV가 가정에까지 보급되기 시작했다[1]. 과거에도 그랬던 것처럼, 3차원 영상 산업은 순간적인 인기로 그칠 것이라 비관적으로 전망했던 전문가들도 현재의 추세를 봤을 때는 3차원 영상 산업의 열기가 당분간 지속될 것으로 전망을 돌리고 있다.

3차원 비디오 시스템은 다시점 색상영상과 이에 상응하는 깊이 영상을 포함하는 3차원 비디오의 획득/처리/전송/재생에 대한 모든 과정을 포함하고 있다[2]. 3차원 비디오는 스테레오 카메라, 깊이 카메라, 다시점 카메라 등 다양한 저작 도구들을 이용하여 직접 얻을 수 있으며, 특히, 깊이 카메라를 이용할 경우 깊이 영상을 직접 획득할 수 있다. 하지만, 스테레오 카메라 혹은 다시점 카

메라의 경우나 기존에 제작된 2차원 영상의 경우, 깊이 영상을 직접적으로 획득할 수 없기 때문에 촬영된 장면으로부터 3차원 정보를 계산적으로 추정하여 깊이 영상을 획득해야 한다.

다시점 영상을 이용해서 깊이 영상을 획득하기 위한 방법으로는 스테레오 정합 방법이 일반적으로 가장 많이 사용되고 있다[3-5]. 이 방법은 비교적 높은 정확도의 깊이 영상을 제공할 수 있지만, 매 화면마다 독립적으로 깊이 영상을 탐색하기 때문에 깊이 영상의 시간적 상관도가 낮은 단점이 있다. 시간적 상관도 저하 문제는 깊이 영상을 이용하여 가상 시점의 영상을 합성했을 때, 장면이 떨리게 되어, 시청자에게 시각적인 불편함을 유발한다. 또한, 깊이 영상 자체의 부호화 과정에서 시간적 예측 효과를 저하시키기 때문에 깊이 영상의 부호화 효율을 저하시킨다[6].

본 논문에서는 깊이 영상 탐색 과정에서 발생하는 시간적 상관도 저하 문제를 해결하여 합성 영상의 화질과 깊이 영상의 부호화 효율을 동시에 높이는 방법을 제안한다. 제안하는 방법은 스테레오 정합 방법을 통해 얻은 깊이 영상에 대해 색상 영상의 경계 정보와 색상 정보를 깊이 영상의 필터링에 사용하는 결합형 양방향 필터를 사용한다. 또한, 움직임이 있는 영역에 대해서는 필터링 오차가 발생할 수 있기 때문에 외곽 오차 제거 기술을 이용해서 필터링을 선별적으로 적용한다.

\* 교신저자(Corresponding Author) : 호요성, 주소 : 광주광역시 북구 오룡동 광주과학기술원(500-712), 전화 : 062)715-2211, E-mail : hoyo@gist.ac.kr

\* 광주과학기술원 실감방송연구센터

\*\* 광주과학기술원 정보통신공학부

(E-mail : sblee@gist.ac.kr)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음(NIPA-2011-(C1090-1011-0003)).

본 논문은 아래와 같이 구성된다. 2장에서는 깊이 영상 탐색을 위한 스테레오 정합 방법에 대해 소개하고, 3장에서는 제안하는 깊이 영상 필터링 방법을 설명한다. 4장에서는 제안한 방법을 이용해 실험 결과를 분석하고, 5장에서 결론을 맺음으로 본 논문을 마무리 한다.

## 2. 깊이 영상 탐색을 위한 스테레오 정합

스테레오 카메라를 이용한 변위 계산 방법에서 변위와 깊이의 상관관계는 다음과 같은 방법으로 유도할 수 있다. 스테레오 카메라는 정확하게 수평이 맞도록 정렬이 되어 있고, 또한 카메라의 광축이 정확하게 평행을 이룬다고 가정한다. 이러한 카메라 구성에서 멀리 있는 물체는 좌영상과 우영상에서 시차가 작게 발생하는 반면, 가까이 있는 물체는 시차가 크게 나타난다. 이러한 원리를 이용해서 기준 시점의 모든 화소가 참조 시점의 어느 위치에 존재하는지 탐색하면 변위(disparity)를 얻을 수 있고, 변위를 이용하면 실제 깊이 정보를 계산할 수 있다. 그림 1은 변위 계산의 기본 원리를 도식화한 것이다.

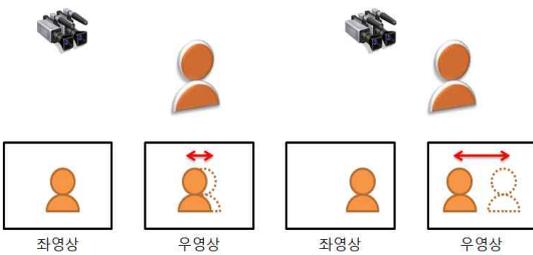


그림 1. 변위 계산의 기본 원리

그림 2는 변위와 깊이의 상관관계를 보다 자세히 나타낸 것이다. 우선, 두 대의 카메라가 각각  $C_l$ ,  $C_r$ 에 위치하고, 두 카메라의 광축이  $z$ 축 방향으로 평행하다고 가정을 한다. 이와 같은 환경에

서 3차원 공간상에서의 한 점  $P$ 는 좌측 카메라의 영상 평면에서  $(x_l, y)$ 에 사상(mapping)되고, 우측 카메라의 영상 평면에서  $(x_r, y)$ 에 사상된다. 그렇다면,  $x_l$ 과  $x_r$ 의 절대값 차이는 변위  $d$ 가 되고, 삼각 비례식을 통해 실제 깊이 정보  $Z$ 를 계산할 수 있게 된다.

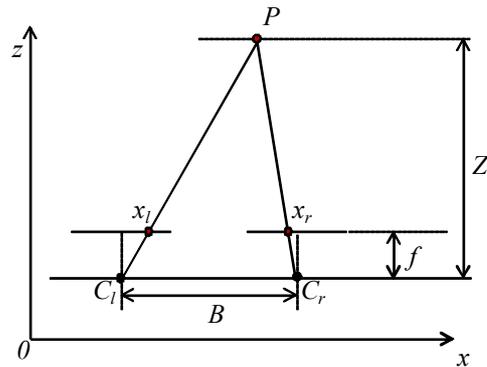


그림 2. 변위-깊이의 상관관계

인간 시각 체계(human visual system)는 양쪽 눈으로부터 얻은 영상의 시차로 물체의 깊이 정보를 인지하는 것으로 알려져 있다. 스테레오 정합(stereo matching) 기술은 이러한 인간 시각 체계를 컴퓨터 시뮬레이션에 차용한 예로써, 현재까지도 컴퓨터 비전 분야에서 활발히 연구 중이다 [3]. 스테레오 영상에서 객체의 변위값을 구하기 위해서는 두 영상에 모두 존재하는 동일한 객체의 화소 위치를 찾아야 한다. 이러한 과정을 대응점 탐색이라고 하며 그림 3과 같이 좌영상에 존재하는 한 물체의 대응점을 찾기 위해 우영상에서 수평의 화소들에 대하여 실제 대응되는 점을 찾게 된다.

변위를 계산하기 위해서는 정합 함수를 이용하여 정합 비용을 계산하며, 정합 비용이 가장 작은 값을 가질 때의 변위를 취하도록 한다. 정합 함수는 두 가지 가정을 포함하고 있는데 첫 번째는 두 시점에서 변위만큼 떨어진 화소는 서로 유사해

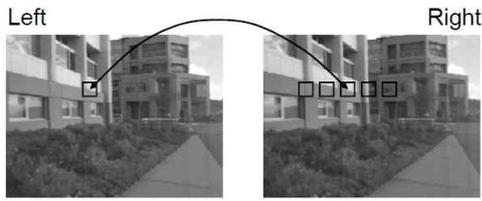


그림 3. 대응점 탐색 예시

야 한다는 것이며, 두 번째는 주변의 화소들은 비슷한 변위를 가져야 한다는 것이다. 이 두 가지 가정은 정합 함수에서 데이터 항(data term)과 평활화 항(smoothness term)으로 표현된다. 정합 함수는 다음과 같이 정의된다.

$$D(x, y) = \arg \min_d \{E(x, y, d)\} \quad (1)$$

$$E(x, y, d) = E_{data}(x, y, d) + E_{smooth}(x, y, d) \quad (2)$$

여기서  $D(x,y)$ 는 화소  $(x,y)$ 에서의 변위를 말하고, 정합 함수  $E(x,y,d)$ 는 데이터 항  $E_{data}(x,y,d)$ 와 평활화 항  $E_{smooth}(x,y,d)$ 로 구성된다.

스테레오 정합 방법은 변위 계산 방법에 따라 크게 지역적인 것과 전역적인 방법으로 나눌 수 있다. 일반적으로 지역적 방법은 계산 속도가 빠른 반면, 화소별로 각각 대응점을 찾기 때문에 폐색 (occlusion) 영역, 반복되는 텍스처 혹은 텍스처가 없는 영역, 빛의 반사가 서로 다른 시점에서 동일하지 않은 평면 등 정보가 모호한 영역에서는 성능이 저하되는 단점이 있다. 전역적 방법은 주변의 정보를 같이 활용하여 영상 전체의 오차를 최소화하기 때문에 보다 정확한 결과를 얻을 수 있지만, 계산 시간은 더 많이 소요된다. 대표적인 전역적 방법으로는 마르코프 랜덤 필드 (Markov Random Field, MRF) 모형을 들 수 있는데 [7], 이 방법은 입력 영상을 이용하여 필드를 구성하고 최소 에너지를 구하여 변위를 계산한다. 따라서, 필드를 구성하는 방법과 최소 에너지 함수를 정의

하는 것이 알고리즘의 성능을 좌우한다. 최근에는 전역적 방법이 지역적 방법에 비해 상대적으로 높은 성능을 보이면서 많은 연구들이 이 접근 방법을 따르고 있다 [8]-[11]. 그림 4는 스테레오 정합 기술을 통해 획득한 변위 영상을 나타내고 있다. 그림 4에서 얻은 변위 영상은 전역적 방법 중 하나인 신뢰 전파 (belief propagation) 기술을 이용해서 얻은 결과이다 [10]. 그림 4(b)에서 알 수 있듯이, 스테레오 정합 기술을 이용하면 객체의 깊이 정보를 효과적으로 얻을 수 있다.

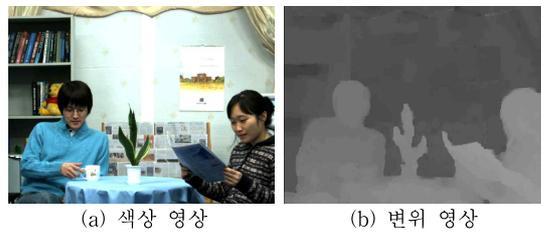


그림 4. 스테레오 정합 기술로 얻은 변위 영상

### 3. 깊이 영상의 후처리 필터링 방법

#### 3.1 결합형 양방향 필터

스테레오 정합 기술을 이용하면 3차원 장면의 깊이 정보를 획득할 수 있지만, 그림 5에서 알 수 있듯이 색상 영상과 변위 영상의 객체 경계가 맞지 않는 문제점이 발생한다. 이러한 문제점은 전역적 방법을 사용했을 때 자주 발생하는 것으로 영상을 합성하는 과정에서 오차를 발생시킬 뿐만

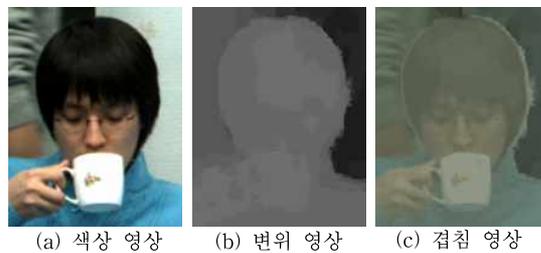


그림 5. 색상 영상과 변위 영상의 경계 불일치

아니라, 다양한 3차원 응용 분야에 깊이 정보를 사용할 수 없게 한다.

최근, 이와 같은 문제점을 해결하기 위해서 획득한 변위 영상에 대해 후처리 필터링을 적용하는 방법들이 제안됐다[12,13]. 후처리 필터링은 기본적으로 양방향 필터(bilateral filter)로부터 출발한다. 기존의 양방향 필터는 한 장의 색상 영상에 적용되는 저역 필터로써, 인접 화소와의 거리차, 색상차 각각에 대한 두 개의 가우시안 (Gaussian) 함수를 사용한다. 새롭게 제안된 결합형 양방향 필터(joint bilateral filter)에서는 두 개의 가우시안 함수에 추가적으로 인접 화소와의 변위차에 대한 항을 사용한다. 그림 6은 변위 영상에 후처리 필터링을 적용한 결과를 나타낸다. 그림 6(b)과 그림 6(d)에서 알 수 있듯이, 필터링을 적용하기 전 변위 영상에 비해 객체의 경계가 상당 부분 일치하는 것을 확인할 수 있다.

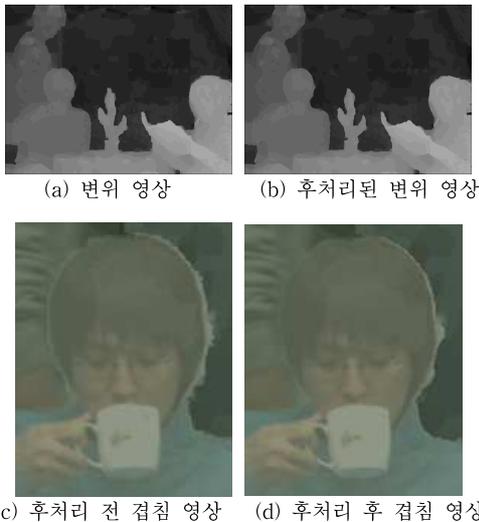


그림 6. 결합형 양방향 필터링 결과

### 3.2 깊이 영상의 시간적 상관도 저하 문제

시간적 상관도 저하 문제는 매 화면마다 독립적으로 깊이 영상을 탐색하면서 깊이값이 시간적

으로 변동폭이 심하게 변하는 문제를 말한다. 움직임이 없는 객체의 경우, 매 화면마다 동일한 깊이값을 가져야 하지만 깊이 영상 탐색 방법의 오차로 인해 화면마다 다른 깊이값을 가지게 된다. 그로 인해, 깊이 영상을 이용하여 중간영상을 합성했을 때, 장면이 떨리는 깜빡임 (flickering) 현상을 발생시켜 시청자에게 시각적 불편함을 유발한다. 또한, 깊이 영상 부호화에서 시간적 예측 효과를 저하시키기 때문에 부호화 효율을 떨어뜨린다. 그림 7은 시간적 상관도가 낮은 깊이 영상의 연속된 세 화면을 나타낸다.



그림 7. 깊이 영상의 시간적 상관도 저하 문제

깊이값의 변동량을 정량적으로 조사하기 위해서, 우리는 그림 8(a)에 표시된 사각형 내의 평균 깊이값을 측정했다. 이 사각형 내의 영역은 100 화면 동안 움직임이 없는 영역이다. 그림 8(b)에서 알 수 있듯이, 기존의 방법으로 얻은 깊이 영상의 100 화면 동안의 평균 깊이값이 심하게 변하는 것을 확인할 수 있다.

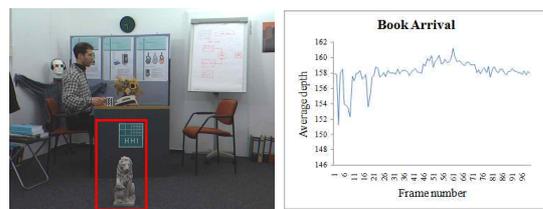


그림 8. 움직이지 않는 영역에 대한 깊이값의 변동폭

### 3.3 결합형 양방향 필터를 이용한 시간적 상관도 향상 방법

시간적 상관도 저하 문제를 해결하기 위한 기

존의 알고리즘들은 대부분이 깊이 영상을 탐색하는 과정에서 움직임 예측을 통해 이전과 이후 장면에서의 깊이 값을 참조하는 방법으로 시간적 상관도를 향상시킨다[14,15]. 하지만, 이러한 기존 기술들은 움직임 예측의 성능에 따라 그 결과가 상당히 의존적이며, 움직임 예측 과정에서 상당한 계산 시간이 소요된다.

본 논문에서는 기존 기술에서 사용한 양방향 필터를 시간축으로 확장하는 깊이 영상 후처리 필터링 방법을 제시한다. 제안하는 방법은 필터 적용 범위를 3차원으로 확장해서 사용한다. 시간축으로 확장된 결합형 양방향 필터는 다음과 같이 정의된다.

$$D(x, y, t) = \arg \min_{d \in d_p} \frac{\sum_{u \in u_p} \sum_{v \in v_p} \sum_{w \in w_p} W(u, v, w) \cdot C(u, v, w, d)}{\sum_{u \in u_p} \sum_{v \in v_p} \sum_{w \in w_p} W(u, v, w)} \quad (3)$$

위 식에서 각 변수는  $p=(x,y)$ ,  $d_p=\{D(x-1,y,t), D(x+1,y,t), D(x,y-1,t), D(x,y+1,t), D(x,y,t-1), D(x,y,t+1)\}$ ,  $u_p=\{x-r, \dots, x+r\}$ ,  $v_p=\{y-r, \dots, y+r\}$ ,  $w_p=\{t-r, \dots, t+r\}$ 와 같고  $D(x,y,t)$ 는 변위값을 나타낸다. 위 식에서 양방향 필터의 가중값인  $W(u, v, w)$ 와 정합 비용  $C(u, v, w, d)$ 는 다음과 같이 정의된다.

$$W(u, v, w) = \exp \left\{ - \frac{\|I(x, y, t), I(u, v, w)\|^2}{2\sigma_r^2} \right\} \cdot \exp \left\{ - \frac{(x-u)^2 + (y-v)^2 + (t-w)^2}{2r^2} \right\} \quad (4)$$

$$C(u, v, w, d) = \min(\lambda L, |D(u, v, w) - d|) \quad (5)$$

여기서  $\lambda$ 는 정합 오차  $C(u, v, w, d)$ 의 상한을 결정하기 위한 상수를 나타내고  $L$ 은 스테레오 정합에 사용된 변위값의 범위를 나타낸다. 그리고  $I(x,y,t)$ 는 색상 영상의 화소값을 나타낸다.

위와 같은 결합형 양방향 필터를 시간축으로 확장해서 깊이 영상에 적용하게 되면 깊이 영상

내의 대부분의 영역의 시간적 상관도를 향상시킬 수 있지만 움직임이 급격한 객체에 대해서는 필터링 이후에 오히려 깊이값의 오차가 발생할 수 있다. 따라서, 본 논문에서는 이를 해결하기 위해 시간축으로 발생하는 외곽 오차를 제거하기 위한 방법을 추가적으로 적용한다.

제안하는 방법은 양방향 필터의 3차원 윈도우 내의 이전 화면과 이후 화면에서  $(x,y)$  위치의 깊이값을 조사한다. 이 깊이값들 가운데 현재 화면과의 차이가 한계값 이상이고 화소값 차이 역시 또다른 한계값 이상일 경우에는 외곽 오차가 포함된 화면이라 판단하여 이 화면을 제거한 후 필터링을 수행한다. 이를 수식으로 나타내면 다음과 같다.

$$w_{outlier\_reduction} = \left\{ w_p \mid \begin{aligned} &|I(x, y, t) - I(x, y, w_p)| < 2\lambda L \\ &|D(x, y, t) - D(x, y, w_p)| < \lambda L \end{aligned} \right\} \quad (6)$$

#### 4. 실험 결과 및 분석

제안하는 방법의 성능을 평가하기 위해, 본 논문에서는 MPEG의 3차원 비디오 부호화 그룹에서 테스트 영상으로 사용 중인 “Cafe”, “Dancer” 영상을 사용하였다. 영상의 해상도는 두 영상 모두 1920x1080이며 “Dancer” 영상의 경우에는 컴퓨터 그래픽스 저작 도구를 위해 제작됐으며 원본 깊이 영상이 존재한다.

깊이 정보의 정확도 평가를 위해서, 원본 깊이 영상이 존재하는 “Dancer” 영상에 대해서 실험을 수행했다. 깊이 영상에 후처리 필터링을 적용하기 위해서 MPEG 실험을 위해 개발된 Depth Estimation Reference Software (DERS) 5.0을 사용해서 깊이 영상을 획득하고 필터링을 적용해서 원본 깊이 영상과의 비교를 통해서 정확도를 PSNR 값으로 나타냈다. 표 1은 깊이 영상의 정확도 비교

표 1. 깊이 영상의 정확도 비교

시점	깊이 영상 정확도 (dB)		
	DERS	기존 필터링 방법	제안하는 방법
5	37.5605	39.6492	40.2697

결과를 나타낸다. 표 1에서 알 수 있듯이, 기존의 2차원 필터링 방법은 DERS에 비해 약 2 dB 정도 깊이 영상의 정확도를 증가시켰고, 제안하는 방법은 약 2.7 dB 정도 정확도를 증가시켰다.

제안하는 방법이 깊이 영상의 부호화에 미치는 영향을 평가하기 위해서 “Cafe” 영상을 이용해서 깊이 영상을 획득하고 후처리 필터링을 적용한 후에 H.264 부호화기를 이용하여 부호화하고 이를 비교하였다. 화면 수는 100 화면으로 정하였고, GOP 구조는 IPPP... 구조를 이용하였다.

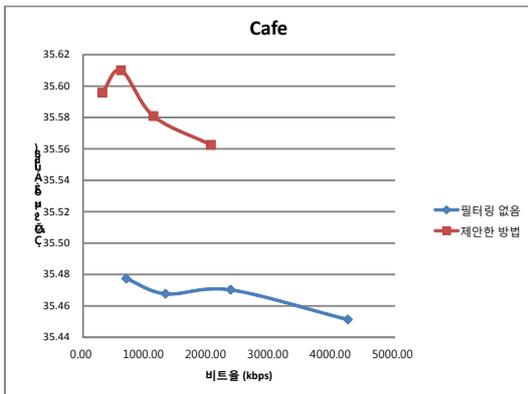


그림 9. 합성 화질-비트율 곡선

표 2. 깊이 영상의 부호화 비교

QP	합성영상 화질 (dB)		비트율 (kbps)	
	필터링 없음	제안한 방법	필터링 없음	제안한 방법
22	35.45	35.56	4250.85	2050.24
25	35.47	35.58	2368.83	1128.36
28	35.47	35.61	1324.40	609.33
31	35.48	35.60	694.52	310.05

표 2와 그림 9는 깊이 영상을 부호화 한 결과를 나타낸다. 부호화 결과를 통해 제안하는 방법이 필터링을 적용하지 않은 결과에 비해 평균 약 0.12 dB의 화질 향상 혹은 53.37%의 비트율을 감소시킨다는 것을 확인하였다.

제안한 방법의 깊이 영상 부호화에 대한 보다 자세한 분석을 위해, 우리는 양자화 매개변수 22에서 깊이 영상 부호화의 부호화 모드 분포를 조사했다. 그림 10은 앞의 깊이 영상 부호화 실험에서의 부호화 모드 정보를 나타낸다. 그래프에서 알 수 있듯이, 제안한 방법을 통해 깊이 영상의 시간적 상관도가 향상됨에 따라, 깊이 영상 부호화에서 inter 예측이 급격히 증가했고, 이로 인해 부호화에서 비트율이 상당히 감소한 것을 확인할 수 있었다.

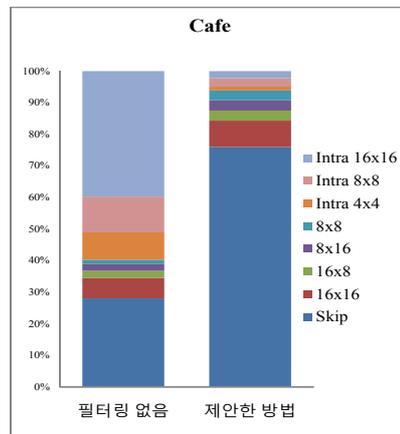


그림 10. 깊이 영상 부호화의 부호화 모드 정보

### 5. 결 론

본 논문에서는 깊이 영상의 시간적 상관도 저하 문제를 해결하기 위해 깊이 영상에 후처리 필터링 기술을 적용하여 합성 영상의 화질과 깊이 영상의 부호화 효율을 동시에 높이는 방법을 제안했다. 제안하는 방법은 시간축으로 확장된 결합형

양방향 필터를 사용했고, 움직임이 있는 영역에서 발생하는 필터링 오차를 줄이기 위해 외곽 오차 제거 기술을 적용했다. 본 논문에서 제안한 방법을 적용했을 때, 깊이 영상 부호화에서 평균 약 0.12 dB의 합성 영상의 화질 향상을 가져왔으며 약 53.37%의 비트율 감소 효과를 볼 수 있었다.

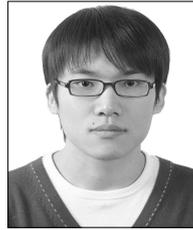
### 참 고 문 헌

- [1] “제임스 캐머런 ‘상상력, 기술의 新 르네상스’,” <http://www.mt.co.kr/view/mtview.php?type=1&no=2010051308483641428&outlink=1>.
- [2] ISO/IEC JTC1/SC29/WG11, “Applications and Requirements on FTV,” N9466, Oct. 2007.
- [3] D. Sharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” in Proceedings of IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, 2001.
- [4] I.Y. Shin and Y.S. Ho, “Disparity Estimation at Virtual Viewpoint for Real-time Intermediate View Generation,” in Proceedings of Systems and Applications 2011 (3DSA 2011), paper S6-2, pp. 195-198, 2011.
- [5] W.S. Jang and Y.S. Ho, “Disparity Map refinement using occlusion handling for 3D scene reconstruction,” in Proceedings of International Conference on Embedded Systems and Intelligent Technology (ICESIT), pp. 213-216, Feb. 2011.
- [6] 이상범, 이천, 호요성, “3차원 영상 생성을 위한 깊이맵 추정 및 중간시점 영상합성 방,” 한국통신학회 논문지, 제34권, 제10호, pp. 1070-1075, 2009.
- [7] P. Perez, “Markov random fields and images,” *CWI Q.*, Vol.11, No.4, pp. 413-437, 1998.
- [8] M. Bleyer and M. Gelautz, “Graph-based surface reconstruction from stereo pairs using image segmentation,” in Proceedings of SPIE EI, Vol.5665, pp. 288-299, 2005.
- [9] Y. Deng, Q. Yang, X. Lin, and X. Tang, “A symmetric patch-based correspondence model for occlusion handling,” in Proceedings of ICCV, pp. 1316-1322, 2005.
- [10] P.F. Felzenszwalb and D.P. Huttenlocher, “Efficient belief propagation for early vision,” in Proceedings of CVPR, pp. 261-268, 2004.
- [11] J. Sun, Y. Li, S. Kang, and H. Shum, “Symmetric stereo matching for occlusion handling,” in Proceedings of CVPR, pp. 399-406, 2005.
- [12] Q. Yang, L. Wang, and N. Ahuja, “A constant-space belief propagation algorithm for stereo matching,” in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.
- [13] P. Lai, D. Tian, and P. Lopez, “Depth map processing with iterative joint multilateral filtering,” in Proceedings of Picture Coding Symposium, pp. 9-12, 2010.
- [14] H. Tao, H.S. Sawhney, and R. Kumar, “Dynamic Depth Recovery from Multiple Synchronized Video Streams,” in Proceedings of CVPR, pp. 118-124, 2001.
- [15] G. Zhang, J. Jia, T. Wong, and H. Bao, “Consistent Depth Maps Recovery from a Video Sequence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.31, No.6, pp. 974-988, 2009.



호 요 성

- 1981년 서울대학교 전자공학과(학사)
- 1983년 서울대학교 전자공학과(석사)
- 1989년 Univ. of California, Santa Barbara, Department of Electrical and Computer Engineering(박사)
- 1983년~1995년 한국전자통신연구소 선임연구원
- 1990년~1993년 미국 Philips 연구소, Senior Research Member
- 1995년~현재 광주과학기술원 정보통신공학과 교수
- 관심분야 : 디지털 신호처리, 영상신호 처리 및 압축, 멀티미디어 시스템, 디지털 TV와 고선명 TV, MPEG 표준, 3차원 TV, 실감 방송



이 상 범

- 2004년 경북대학교 전자전기공학부(학사)
- 2006년 광주과학기술원 정보통신공학과(석사)
- 2006년~현재 광주과학기술원 정보통신공학과 박사과정
- 관심분야 : 디지털 영상처리, 3차원 TV, 깊이 영상 획득, 3차원 영상 부호화, 실감 방송